Joint Virtual Reality Conference of ICAT - EGVE - EuroVR (2012)
R. Boulic, C. Cruz Neira, K. Kiyokawa, and D. Roberts (Editors)

An Empiric Evaluation of Confirmation Methods for Optical See-Through Head-Mounted Display Calibration

Patrick Maier[†1], Arindam Dey[‡2], Christian A.L. Waechter[§1], Christian Sandor[¶2], Marcus Tönnis[‖1], Gudrun Klinker[**1]

[1]Fachgebiet Augmented Reality (FAR), Technische Universität München, Fakultät für Informatik, Germany
[2]Magic Vision Lab, University of South Australia, Adelaide, Australia

(a) Keyboard (b) Hand-held (c) Voice (d) Waiting

Figure 1: *Compared confirmation methods: (a) Keyboard, (b) Hand-held, (c) Voice, and (d) Waiting. The Waiting method was the most accurate in data collection for OSTHMD calibration. Averaging over time frames further improved the calibration.*

Abstract

The calibration of optical see-through head-mounted displays (OSTHMDs) is an important fundament for correct object alignment in augmented reality. Any calibration process for OSTHMDs requires users to align 2D points in screen space with 3D points and to confirm each alignment. In this paper, we investigate how different confirmation methods affect calibration quality. By an empiric evaluation, we compared four confirmation methods: Keyboard, Hand-held, Voice, and Waiting. We let users calibrate with a video see-through head-mounted display. This way, we were able to record videos of the alignments in parallel. Later image processing provided baseline alignments for comparison against the user generated ones. Our results provide design constraints for future calibration procedures. The Waiting method, designed to reduce head motion during confirmation, showed a significantly higher accuracy than all other methods. Averaging alignments over a time frame improved the accuracy of all methods further more. We validated our results by numerically comparing the user generated projection matrices with calculated ground truth projection matrices. The findings were also observed by several calibration procedures performed with an OSTHMD.

Categories and Subject Descriptors (according to ACM CCS): [H.5.1]: Multimedia Information Systems—Artificial, augmented, and virtual realities [H.1.2]: User/Machine Systems—Human Factors;

1. Introduction

A big challenge in Augmented Reality (AR) is to achieve a seamless integration of virtual objects into the real world. Irrespective of the display technology used, display calibration is always required in an AR system. To achieve accurate calibration in video see-through head mounted displays (VSTHMD), computer vision algorithms are used to

† e-mail: maierp@in.tum.de
‡ e-mail: arindam.dey@unisa.edu.au
§ e-mail: waechter@in.tum.de
¶ e-mail: christian@sandor.com
‖ e-mail: toennis@in.tum.de
** e-mail: klinker@in.tum.de

DOI: 10.2312/EGVE/JVRC12/073-080

find 2D-3D correspondences, requiring a minimum of user interaction. In contrast, optical see-through head mounted displays (OSTHMD) require a higher level of user interaction for calibration. Here the user has to repeatedly align a 2D and a corresponding 3D point manually to define proper correspondences. Such correspondence data is used in calibration mechanisms for OSTHMD such as the Single Point Active Alignment Method (SPAAM) [TGN02] and the Display Relative Calibration (DRC) [OZTX04]. While most of the parameters required for SPAAM calibration are captured by human users, DRC is based on a two phase calibration method where only in the second phase (using option 4 of the DRC paper [OZTX04]) user interaction is required to calibrate the display system for a specific placement on the user's head. SPAAM is therefore more vulnerable to the human error during the calibration process than DRC.

After the alignment of a 2D point on the display with a 3D point in the real world, users usually confirm the correspondence by pressing a key on the keyboard. We assert this confirmation method plays a major role on the human performance during the process. The confirmation action requires some degree of hand coordination forcing a misalignment of the 2D point with the 3D point, increasing the probability to capture inaccurate data. We expect that creating an interface that minimizes the required locomotion during confirmation could lead to a more precise acquisition of calibration data, increasing the augmentation accuracy of the system.

In this paper, we evaluate three confirmation methods, *Hand-held* (pressing a hand-held button), *Voice* (verbally reporting) and *Waiting* (detecting nearly no movement of the head for at least 0.5 second) with the most often used method *Keyboard* (pressing a key on the keyboard) in a user study. The study aims on investigating the influence of the confirmation methods on the quality of the measurements.

To compare the different confirmation methods, we had to measure the misalignment of the 2D and 3D point by the user. Collecting quantitative data was enabled by using a VSTHMD. A computer vision algorithm applied to the video stream calculates the misalignment of the user specified 2D points to the corresponding 3D points in pixels. Procedures as done by Navab et al. [NZGC02] where the calibration quality of a OSTHMD is evaluated in a quantitative manner on-line, can not be applied here: further correspondence point selections which are necessary to evaluate the calibration on-line, additionally induce new user-generated errors, in our case distorting the results.

Using a VSTHMD instead of an OSTHMD appears feasible because the focus lies on the minimization of head motion during the confirmation process; and head motion is mainly dependent on the confirmation method not on the type of HMD. Confirmation methods which reduce the head motion during confirmation will also reduce the error for any type of HMD. In general, the alignment error on both types of HMD mainly consists of an accuracy error done by the user (even if the user would not shiver or had enough time to do the alignment) and the error induced by the confirmation method. Using an OSTHMD, an additional error is induced by the eye point which is not fixed according to the HMD. By minimizing the error induced by the confirmation method, the error for both types of HMD is reduced. Because of this, we argue that using a VSTHMD instead of a OSTHMD can be used to analyze the errors induced by the confirmation methods. A subjective expert test at the end of section 6 also shows the same trends by using an OSTHMD.

The results of the study show that the *Waiting* method outperformed all three other confirmation methods including the currently most-used keyboard-based method (see Fig. 4).

We furthermore investigated the application of different time frames to average collected data to yield further improved results. Besides generally smoothing data over a set of samples, we made the observation that test participants tended to slightly shake their head or directly proceeded to the next point, right after an acknowledgment. Time frames of varying length ($0.1sec$ to $2sec$) were inspected, all ending at the time of confirmation $[t]$. In addition, the symmetric time frame $[t-0.25sec;t+0.25sec]$ was inspected, verifying our previous assumption that the error in this time frame was too high, because of movements caused by the input method and an already ongoing movement to the next correspondence point. To analyze the different time frames, we recorded the tracking data during the whole alignment and confirmation process.

We found that the time frame of $[t-0.6sec,t]$ resulted in the most accurate alignment among the above mentioned time frames. The current approach of calculating the residual error just at the point of acknowledgment $[t]$ in comparison resulted in the worst accuracy. Time windows that end before the time of confirmation in the form of $[t-a,t-b]$ will be analyzed in the future as written in section 7.

In the next section, we review previous literature relevant to this work. In the section covering the experiment (Sec. 3), we provide details about the three acknowledgment methods and their implementation (Sec. 3.1). Next we present the results of the user study (Sec. 4) and discuss them in Sec. 5. Before we conclude with the final discussion and directing towards the conclusions and future work in Sec. 7, we validate the results of the experiment in Sec. 6.

2. Background

Since the publication of calibration methods for HMDs, work on improving the calibration mainly investigated numerical aspects, focusing on computation of perspective and transformational parameters, minimization of errors and on target placement. Human factors have been investigated to a lesser degree, mostly as a side effect of target placement and aiming.

2.1. Numerical Aspects

In the first publication of the SPAAM algorithm Tuceryan and Navab [TGN02] noticed that the more of the tracker volume is covered by moving the user's head, the more possible systematic errors in the tracker measurements will be taken into account in the optimization process. They encourage the user to move the head around as much as possible while the calibration process is executed, but not during confirmation, but relativize this argument due to tracker issues in their setup. The tracker also induces lag in the tracker data at the point of collection. If the button is clicked too quickly, the tracker data read may not correspond to where the user's head is. Newer tracking systems may have a lower lag, but the question remains when and how to capture the data.

Axholt deeply investigated distribution of correspondence points [ASP*10] and found that a wider distribution in depth is an influential parameter towards good calibration precision. The addition of more correspondence points can lead towards good calibration precision, but consumes more time and (depending on the calibration exercise) can potentially increase the precision, but also exhausts the user to the point where alignments are not as good as they could be. It therefore appears more useful to consider how to distribute correspondence points in depth.

Further work by Axholt examined this distribution of correspondence points in depth [ASO*11]. Three different correspondence point distributions were investigated. Additional simulation results showing improvement factors versus number of correspondence points are presented. Distribution of correspondence points in depth affects the variance of the estimated eye point. Axholt found that almost all parameters of the pinhole camera model depend on the variance of the correspondence point distribution, except for orientation appearing to be more dependent on the number of correspondence points. He also found that a lesser number of correspondence points seem to be necessary when using a random distribution.

2.2. Human Aspects

Demer et al. [DGP91] investigated the unwanted head movements of users with a telescopic spectacles and without. They observed that the subjects always have a certain head movement with different magnitudes depending on magnification and different qualities of vision. Other works investigate aspects of calibration procedures that are related to human behavior. The user's inability to maintain a stable pose [Lip71] is the most relevant parameter when collecting point correspondences as stated by Tuceryan et al. [TGN02] (p8 last paragraph, p12 section 5.1 last sentence).

McGarrity et al. [MGN*01] stated that the user must be factored in when calibrating an HMD. Errors are induced by users because calibration procedures involve manual steps. McGarrity et al. mention factors that might appear to only have small effect, such as facial muscle contractions, e.g., talking, raising the brow, but obstruct calibration quality. In addition they request that simplicity is a must for any calibration algorithm. If users have to do some difficult action, it is likely that they will inject errors into the system. Finally, other factors are listed that may cause errors during calibration and evaluation. For example poor lighting may make target alignment less accurate since the user has difficulties perceiving the target through the darkening display.

Earlier work by Axholt et al. [APE*09] investigated postural stability during the calibration process. They studied alignment performance with head-mounted displays at different levels of azimuth and elevation. While the results show that the viewing direction has a statistically significant effect, the effect can be neglected. In practical settings this effect can be approximated by a circular distribution with standard deviation in sub-angular magnitude.

The already mentioned work by Axholt et al. [ASO*11] also discussed user motion during the calibration process. In the experimental design, the subjects were required to move a lot, probably having induced equipment slippage and thus a higher variance in capabilities to align correspondence points. A lower variance is expected in an other work which plans to do a similar study where the subjects will not move.

Livingston et al. [LEW*06] investigated issues arising after display calibration. Vertical disparity between the images of both eyes can lead to problems for the user in fusing both images into a coherent picture of the 3D world. Their approach adds a final step to the calibration procedure by requesting the user to align nonius lines. With this correction step, some alignment errors can be corrected, but a preceding display calibration is still required.

2.3. Waiting as Input

In 2D user interfaces, waiting as a confirmation method is well established. For example on touch screens in mobile devices, the user touches a spot on the screen and keeps the finger at that place to bring up a menu. In this case the system stores the position of the finger on the screen when the finger touches the display. When the position of the finger does not go further than a predefined radius in a certain time, the system triggers the event. Another well-known system is the tooltip element used in programming for mouse based systems where the tooltip shows up when the pointer is over an element for a specific time (no matter how fast the user is moving the pointer on the element).

Steed [Ste06] provides an elaborate discussion about mechanisms for the confirmation of selections by dwell time. Some applications implement such waiting approaches with dwell time. For instance, Feiner et al. [FMHW97] implemented such an approach in a tourist guide system to confirm labels of points of interest. Focusing such a label for at least a second in the center of the head-worn display

shows further information about the object of concern. Ha and Woo [HW10] used a similar dwell time approach. Here the user has to point on an object, having generated a virtual collision for at least 500*ms* to select the object. The work by Axholt et al. [ASO*11] lets users aim at a correlation point for at least 2*s* after activation of the corresponding cross-hair cursor. All data was collected over this time period. The calibration procedure aimed at collecting the 2*s* interval of correspondence points with low differences in the HMD rotation. If the rotation of the HMD changed more than 0.19° per sample, the data was discarded.

All those waiting methods are context based where the user waits in/on a context (button, object etc.) or waits a given time when the system is in a specific state. Our proposed waiting method is context free. The algorithm has to determine everywhere and to every time when the user stops moving. A detailed description of this method can be found in section 3.1.

3. Experiment

The primary design goal of this experiment aimed at evaluating the different acknowledgment methods for the OST-HMD calibration. The experiment investigated how different confirmation methods influence the accuracy of the input data (correspondence points) for the calibration process.

3.1. Acknowledgement Methods

The user has to align a 3D point of the real world at an appropriate depth with a cross hair displayed on the VST-HMD. To acknowledge the correct overlay of the correspondence points, the user has to use one of the four methods described below. Once the user thinks that the 3D target point and the 2D display point are successfully aligned, one of the acknowledgment methods has to be used to acknowledge the correspondence points. The currently available acknowledgment method is pressing a key on the keyboard [TGN02]. This method forces the user to move the hand, and in consequence a part of the body, which may result in a misalignment while acknowledging. We have additionally implemented three acknowledgment methods to minimize the complexity of the process and to achieve a higher accuracy.

Hand-held: We have equipped the user with a hand-held button (Fig. 1(b)). The user acknowledges the correspondence points by pressing the button. This process requires only a minimal finger movement decreasing the amount of locomotion. The user can hold the device in a comfortable way which could prevent unwanted movements due to strain.

Voice: Users verbally confirmed by uttering “ok” or “k” when they completed the alignment successfully. In our study we used a wizard-of-oz technique to compensate for possible failures of voice detection systems. This process required no additional device and only required a minimal movement of facial muscles.

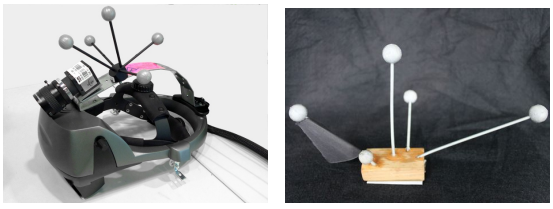
Waiting: The last implemented method requires users to keep their head steady at the point and direction of alignment and wait for at least 0.5 seconds. We therefore call this interaction technique *Waiting* method [MTK10]. We have measured the steadiness of the head by inspecting the normalized vector from the HMD to the target sphere in the HMD coordinate system over the alignment process. To analyze the amount of head movement, the implementation calculates two different values w.r.t. the data of the last 0.5*sec* (as described in more detail elsewhere [MTK10]). The first value is the trajectory length of the normalized vector tip. The second value is the compactness of this trajectory. The compactness is the mean distance of the trajectory points to their centroid. We considered the head to be steady when first, the trajectory length stays below a threshold value of 1.5*cm* and second, when the compactness value is below 0.15*cm* for the last 0.5*sec*. This method requires no additional device to carry and enables users to acknowledge without body movement. In fact, the method forces users to reduce body movement and as a result could lead to a lower aiming error.

3.2. Hypotheses

We postulated the following hypotheses before executing the experiment.
[H1] *Waiting* and *Hand-held* methods will result in the most accurate data. *Voice* and *Keyboard* methods will produce worst accuracy among the methods.
[H2] Averaging over time frames will result in better calibration than taking only the correspondence points at the time of confirmation.

3.3. Experimental Setup

To measure differences in the confirmation methods, we set up a scenario where the users should collect 2D-3D point correspondences. We used an infrared reflecting rigid body as aiming target for the 3D point where the left most silver sphere was clearly indicated (see Figure 2(b)). For the 2D point, nine cross hairs with a surrounding circle were displayed to the user in the HMD one after the other.



(a) (b)

Figure 2: *A VST-HMD was used to calculate the alignment error (a); participants had to align a cross-hair appearing on the HMD with the center of the left most silver sphere in (b).*

To understand human behavior in more detail as well as being able to calculate aiming errors w.r.t. baseline data, we used a OSTHMD (nVisor ST60) converted to a VSTHMD. We therefore mounted a camera on top of the display (see Figure 2(a)). The image of the camera was shown in the display for the dominant eye. The video feed from the camera was also used later to compute the baseline data. Users sat comfortably on a chair, looking in the direction of the 3D target. The 3D target was placed at two different positions, either in $1m$ (near) or $2m$ (far) distance in front of the user at a height of $0.8m$. Users thus could hold their head in a comfortable position when looking at the marker target straight ahead. We used ART[†] for tracking of the HMD and the target. The head movements of the participants were tracked using a rigid body fixed to the HMD. To easier analyze the aiming error, we used a black background to detect marker balls in recorded video feeds via image processing methods in a robust way. The center positions of the marker spheres are estimated using a circle detection algorithm. Estimating the centers of these marker balls in an automatic way provides us with reliable information about the 2D display positions at which the users should have aimed. This baseline data was used to calculate the residual error in comparison to the user acknowledged 2D position.

3.4. Experimental Task

The experimental task in this experiment is similar to the tasks performed to calibrate an OSTHMD and previously described by [TGN02] and [OZTX04]. Participants had to align the cross hair with the silver target sphere by moving the head and body to get to an alignment. Once this alignment is established, they had to use one of our four methods to acknowledge the alignment. Cross hairs were presented one after the other in nine different positions on the display. Users had to repeat this set of eighteen alignments in near and far distance ten times generating 180 correspondence point pairs. Participants were allowed to take a rest after each set. After task completion, participants had to fill out a subjective questionnaire. One set of 18 correspondence points took about 2-3 minutes. Users normally wanted a break after 5 sets of about 2-5 minutes. The experiment took 45 minutes in average.

We have recruited 24 participants from the student and research population of the university, aged between 23 to 53 years (M=28.9, SD=6.05). None of the students had any prior experience with the OSTHMD calibration process; some of them have experienced AR applications before.

3.5. Independent Variables

- **Acknowledgment Method** $\in$ {Keyboard, Hand-held, Voice, Waiting} *between subjects*
  In this experiment, we have randomly distributed the twenty four participants into four different groups having six participants in each group. The participants in one group did the task using only one acknowledgment method described in section 3.1.
- **Correspondence Points:** $\in \{1 \text{ to } 18\}$ *within subjects*
  The experimental setup had two different distance layers – near and far – having nine correspondence points in each layer. The nine correspondence points were distributed in a 3×3 squared orientation on each of the two layers.
- **Repetition** $\in \{1 \text{ to } 10\}$ *within subjects*
  Alignments of the 18 correspondence points were performed in one repetition. The participants performed the experimental task ten times.
- **Time frame:** $\in \{[t-2.0sec,t],...,[t-0.1sec,t],[t]\}$ *within subjects*
  We have recorded the screen-shots of the HMD at 30fps. In post-process, we have calculated the mean residual error in 21 different time frames relative to the point of acknowledgment $[t]$.

3.6. Dependent Variables

We have calculated the residual error in the calibration process as a dependent variable. The whole experiment was based on 4 (methods) $\times$ 21 (time frames) $\times$ 10 (repetitions) $\times$ 6 (participants per method) $\times$ 18 (correspondence points per repetition) = 90720 data points.

4. Results

To statistically analyze the collected data, the residual errors for every time step were calculated. The residual error is the euclidean distance between the red cross hair and the projected center of the silver target sphere in the display image. To calculate the center of the silver target sphere, we used a computer vision algorithm on the recorded screen data. The more users are misaligning cross hair and sphere, the more they are away from the image of the target sphere on the display and the greater the residual error gets. As measurement for the aiming error, we did not only examine residual errors at the time of confirmation t but also the average residual error in the remaining 20 time frames. We have analyzed the effect of the various methods and time frames by a $4\times(21\times10)$ mixed-factorial ANOVA using the statistical package SPSS.

Effect of Acknowledgment Methods: The ANOVA[‡] reported a significant main effect of the acknowledgment methods $F(3,428) = 22.07; p < .001; \eta_p{}^2 = .13$ (Fig. 3). A

[†] http://www.ar-tracking.de

[‡] All of our ANOVA statistics revealed an *Effect Size* in the medium range as denoted by the term $\eta_p{}^2$. It indicates the results have a reasonable practical significance despite the use of six participants per group.

Tukey's HSD post-hoc test revealed that among all methods, the *Waiting* method outperformed and *Keyboard* method being worst. *Waiting* method was significantly ($p < .01$) more accurate and *Keyboard* method was significantly ($p < .001$) less accurate than all other methods. This partly supports our hypothesis *[H1]* for the *Waiting* and *Keyboard* method.

Effect of Time Frames: ANOVA found a significant main effect of time frame $F(1.087, 465.047) = 127.14; p < .001; \eta_p{}^2 = .23$ on the residual error. As the data did not meet the assumption of sphericity, we used Greenhouse-Geisser adjustments[§]. We then performed a pair-wise comparison with Bonferroni adjustments and found that the time frame $[t-0.6sec,t]$ was overall the best, and significantly a better time frame than all time frames except $[t-0.5sec,t]$ and $[t-0.4sec,t]$ time frames. The error calculated at the point of acknowledgment is significantly ($p < .001$) worse than the time frames ranging from $[t-0.9sec,t]$ to $[t-0.1sec,t]$ (see Figure 3) which supports our hypothesis *[H2]*.

There was a significant $TimeFrame \times Method$ interaction effect $F(3.260, 465.047) = 26.42; p < .001; \eta_p{}^2 = .16$. While the difference between the mean residual error was very small for the *Waiting* method across all time frames, for other methods the differences were much bigger. However, for the individual methods the best time frames were different. $[t-0.4sec,t]$ and $[t-0.6sec,t]$ were the best time frames for *Keyboard* and *Hand-held* respectively. For *Voice* the best time frame was $[t-0.8sec,t]$; and $[t-1.7sec,t]$ for *Waiting*. Interestingly, while analyzing the differences between the best time frames of each method with an one-way ANOVA $F(3,4316) = 7.52; p < .001$ and Tukey's post-hoc test, we found the *Waiting* method being significantly better than all other methods *including* the *Keyboard* method (Fig. 4). The *Voice* method was the worst method among these four methods. However, these differences were not significant with the *Keyboard* and the *Hand-held* methods.

We have also investigated the effect of time frames on each individual participant as the OSTHMD calibration is indeed dependent on individual skills. Expectedly, we found the error at $[t]$ was high for any participant in our experiment (supports *[H2]*). $[t-0.6sec,t]$ was the best time frame for 7 participants (1 - *Keyboard*, *Voice*, and *Waiting*; 4 - *Hand-held*) and $[t-0.4sec,t]$ was the best time frame for 3 (*Keyboard*) participants. Similarly, $[t-0.7sec,t]$ was the best time frame for 3 participants (1 each for *Keyboard*, *Hand-held*, and *Voice*). $[t-0.3sec,t]$ was the best time frame for 2 participants (1 each for *Keyboard* and *Waiting*). Interestingly, participants with the *Waiting* method had longer best time frames such as $[t-1.1sec,t]$, $[t-1.7sec,t]$, $[t-1.8sec,t]$, and $[t-2.0sec,t]$. A similar trend is found for the *Voice* method as well with $[t-0.9sec,t]$, $[t-1.0sec,t]$ (2 participants), and $[t-1.5sec,t]$ being the best time frame.

These results are consistent with our hypotheses. The *Waiting* method produces higher accuracy than keyboard based methods *([H1])*. We have found OSTHMD calibration data must be collected in a longer time frame to gain accuracy. Collecting data at the point of confirmation is inaccurate practice *([H2])*.

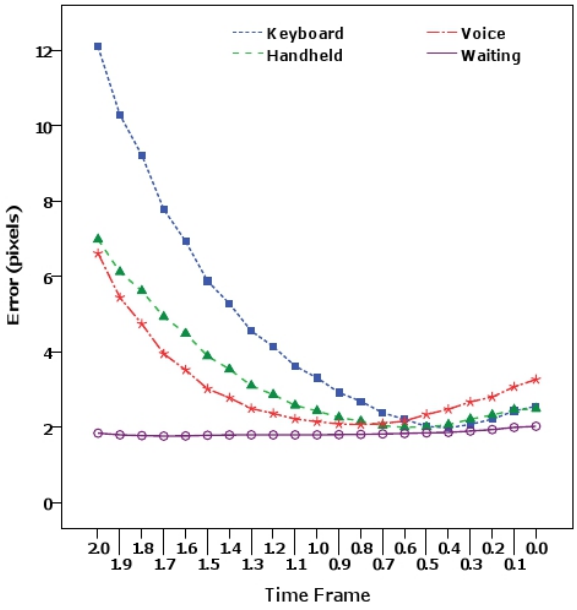


Figure 3: *X-axis represent the time frames $[t-x,t]$. $[t-0.6sec,t]$ was the most accurate time frame. Error was consistently low for Waiting method across the time frames.*

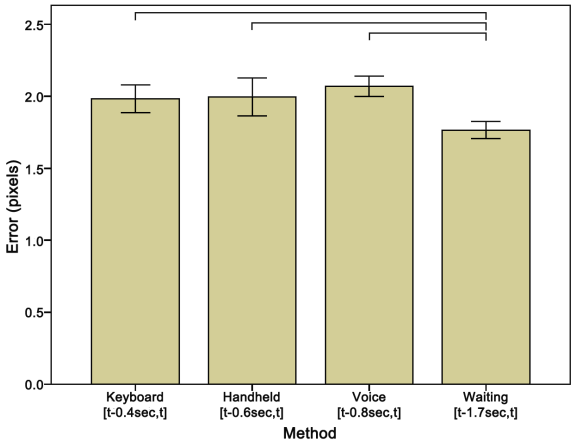


Figure 4: *The best time frame of the Waiting method is significantly better than the best results of all other methods. Whiskers represent ±95% confidence intervals.*

[§] Please see [Bag04] for more details about spericity and the use of Greenhouse-Geisser adjustments.

5. Discussion

Averaging the collected data of time frames results in better calibrations. This behavior can be explained by the movement of the users head while aiming at the target. While aiming, the user oscillates over the target spot with the inability to precisely target this spot. Averaging over those points results in a point that lies closer to the target spot. The longer the user concentrates on the target spot, the better the averaged result should be. But the calibration process should also not be long; otherwise the users get tired and will make mistakes. We let the users do the confirmations in their desired speed not to rush them. This gave a good precondition to find the optimal time frames for the different methods.

We explain the good results of the *Waiting* method in the following way. The users had to concentrate on the target spot trying to keep as still as possible. This reduces the ability to do sloppy confirmations and also calms down the user. Whereas with the other methods, we observed that users kind of rushed over the correspondence points which leads to a higher error than the *Waiting* method.

Participants reported their experience with the calibration methods through a subjective questionnaire. Expectedly, most of the participants reported the heavy weight of the HMD caused them trouble in performing the alignment task. Interestingly, two participants in the *Waiting* method reported that they found aligning the lower points to be harder than the others, as the front heavy HMD was "falling down". We asked participants to report the movement of the head *at the point of* acknowledgment according to their perception. All of the participants in the *Waiting* method used terms like "very little", "minimal", "1mm." etc.; which indicates their high confidence with the task. Whereas in the case of other methods (particularly in *Voice*) participants used terms like "a bit", "1-2 cm.", "some pixels" etc.. Overall, it further indicates the acceptance of the *Waiting* method over other methods. Subjectively, participants did not like the *Voice* method. All of the participants reported a neck fatigue, four participants reported eye fatigue and one participant in the *Voice* method reported the task too stressful.

6. Validation of Results

In the previous study and in its analysis, the value for judging the different methods was the residual error of the 2D display points and the correspondence point of the 3D target on the display. These findings so far leave the question whether the results can be transferred to the quality of the resulting projection matrices.

To validate the gathered results we used a numerical analysis of the user generated calibrations with the ground truth calibration, generated from the computer vision data of the previous experiment. For this numerical analysis, we used the optimal projection matrices that we get from a SPAAM calibration that uses the recorded target 3D points and the calculated corresponding 2D image point from the computer vision algorithm. [OH07] We also calculated the projection matrices of the user data with SPAAM using the cross-hair 2D positions and the corresponding target 3D point. For the 3D positions we generated average positions using the 21 time frames. This way we generated 210 different projection matrices for each user and the optimal projection matrix.

The dimensions of the grid were $2 \times 2 \times 2$ meters with a point every 10 centimeters in each dimension, producing an total number of 9261 discrete values. These points were placed at a distance starting at 2 meters to the HMD such that the farthest point was 4 meters away.

Those 3D points were projected onto the image plane using the 210 user generated projection matrices and calculated for each projection matrix the residual error compared to the points which were projected using the ground truth projection matrix. As a value for comparison we calculated the mean residual error for all visible points on the image plane.

We found a main effect of confirmation methods on the error $F(3,236) = 14.05; p < .001$. A post-hoc test showed the *Waiting* method was significantly better than the *Keyboard* ($p < .001$) and *Voice* ($p = .025$) methods (see Figure 5). Consistent with our previous findings, the *Keyboard* method found to be the significantly worst method and the *Waiting* method to be the best method among the four experimental methods. We also found a main effect of time frame $F(1.36, 320.39) = 54.13; p < .001; \eta_p{}^2 = .19$. Overall, $[t - 0.6sec, t]$ was the best time frame.

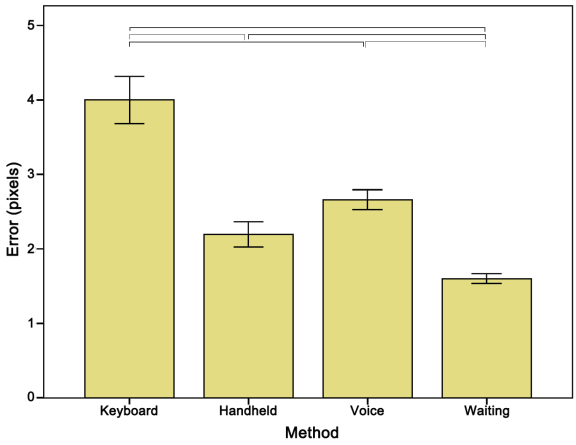


Figure 5: *Our comparison with the ground truth projection matrices confirms that Waiting is the best confirmation method and supports the validity of our approach. Whiskers represent ±95% confidence interval.*

We also observed the same results when comparing the four confirmation methods on an OSTHMD. We had a similar hardware setup as in the experiment except for using an

unmodified OSTHMD this time. Five experienced users did the same calibration procedure using the four different acknowledgment methods as described earlier. Here the users also confirmed 18 correspondence points. We applied the previously calculated best time frames (as shown in Figure 4) for each confirmation method. For each expert user we calculated the four projections and augmented a chessboard like board with its complement pattern. The pattern was colored in a different color for each projection matrix. Each of the four projections was shown in a loop one after the other to investigate user's subjective preference. Each user had to judge the overlay accuracy and had to rank them according to the estimated quality. From user feedback we found that *Waiting* was the best method for everyone. *Voice* was the worst for three users and two other users reported *Keyboard* and *Hand-held* to be the worst methods. Though, five users do not allow for a deeper statistical analysis, *Waiting* is apparently the best method as found in our previous analysis reported in Section 4.

7. Conclusions

Other than numerical improvements in OSTHMD calibration, we focused on the human factor in collecting more accurate input data for the calibration process. We have developed and investigated different confirmation methods for Optical See-Through Head-Mounted Display Calibration. We found that different confirmation methods do have an influence on the quality of the calibration. The *Waiting* method as input for the display calibration resulted in the most accurate correspondence point data. We further found that using averaged correspondence point data of a time frame is always better than only using the correspondence point at the time of confirmation. Each individual method has its own optimal time frame being different from those for other methods. We verified our findings with a numerical analysis of different calibrations. Tries with an OSTHMD also showed the same results.

Implementing the presented contributions for calibration processes for OSTHMD will improve quality of object alignment in AR systems. However, further optimization of user dependent factors might be possible. The analysis of the optimal time frames used time frames which end at the time of confirmation and start at different, earlier points in time. To find the best settings for the time frames, sliding time frames could be an alternative for further examination. With sliding time frames the length of the time frame varies by letting the end point float in time in the interval between the starting point and the time of acknowledgment. By moving the end point to an earlier point in time than the time of acknowledgment, the possibility increases to further suppress effects in locomotion of the user. This is especially the case with the *Voice* method, where an additional delay has to be subtracted because of the processing time for speech recognition.

References

[APE*09] AXHOLT M., PETERSON S., ELLIS S., ET AL.: Visual Alignment Accuracy in Head Mounted Optical See-Through AR Displays: Distribution of Head Orientation Noise. In *Human Factors and Ergonomics Society Annual Meeting Proceedings* (2009), vol. 53, pp. 2024–2028. 3

[ASO*11] AXHOLT M., SKOGLUND M. A., O'CONNELL S. D., COOPER M. D., ELLIS S. R., YNNERMAN A.: Parameter Estimation Variance of the Single Point Active Alignment Method in Optical See-Through Head Mounted Display Calibration. In *Proceedings of the IEEE Virtual Reality Conference* (2011). 3, 4

[ASP*10] AXHOLT M., SKOGLUND M., PETERSON S., COOPER M., SCHN T., GUSTAFSSON F., YNNERMAN A., ELLIS S.: Optical see-through head mounted display direct linear transformation calibration robustness in the presence of user alignment noise. In *Proceedings of the Human Factors and Ergonomics Society 54rd Annual Meeting* (2010). 3

[Bag04] BAGULEY T.: An introduction to sphericity. http://homepages.gold.ac.uk/aphome/spheric.html, 2004. 6

[DGP91] DEMER J., GOLDBERG J., PORTER F.: Effect of telescopic spectacles on head stability in normal and low vision. *Journal of vestibular research 1*, 2 (1991), 109. 3

[FMHW97] FEINER S., MACINTYRE B., HÖLLERER T., WEBSTER A.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal and Ubiquitous Computing 1*, 4 (1997), 208–217. 3

[HW10] HA T., WOO W.: An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *Proceedings of the IEEE Symposium on 3D User Interfaces* (2010), IEEE, pp. 91–98. 4

[LEW*06] LIVINGSTON M., ELLIS S., WHITE S., FEINER S., LEDERER A.: Vertical Vergence Calibration for Augmented Reality Displays. 3

[Lip71] LIPPOLD O.: Physiological tremor. *Scientific American 224* (March 1971), 65–73. 3

[MGN*01] MCGARRITY E., GENC Y., NAVAB N., TUCERYAN M., OWEN C.: Evaluation of calibration for optical see-through augmented reality systems. 3

[MTK10] MAIER P., TÖNNIS M., KLINKER G.: Designing and Comparing Two-Handed Gestures to Confirm Links between User Controlled Objects. In *9th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2010). 4

[NZGC02] NAVAB N., ZOKAI S., GENC Y., COELHO E. M.: An on-line evaluation system for optical see-through augmented reality. In *Presence: Teleoperators and Virtual Environments* (June 2002), vol. vol. 11, pp. pp. 259–276. 2

[OH07] OUELLET J., HEBERT P.: A simple operator for very precise estimation of ellipses. In *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on* (2007), IEEE, pp. 21–28. 7

[OZTX04] OWEN C. B., ZHOU J., TANG A., XIAO F.: Display-relative calibration for optical see-through head-mounted displays. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 70–78. 2, 5

[Ste06] STEED A.: Towards a general model for selection in virtual environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces* (2006), pp. 103–110. 3

[TGN02] TUCERYAN M., GENC Y., NAVAB N.: Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. 259–276. 2, 3, 4, 5